DOCUMENT RESUME

ED 385 587 TM 024 025

AUTHOR	Adams, Richard; And Others
TITLE	Item Difficulty Adjustment Study: GRE Verbal Discretes. GRE Board Professional Report No. 89-04P.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Graduate Record Examinations Board, Princeton, N.J.
REPORT NO	ETS-RR-92-79
PUB DATE	Apr 93
NOTE	34p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	College Entrance Examinations; *Cost Effectiveness; *Difficulty Level; Distractors (Tests); Higher Education; Test Construction; Test Format; *Test Items; *Verbal Tests
IDENTIFIERS	Analogies; Antonyms; *Graduate Record Examinations; *Revision Processes; Sentence Completion Method

ABSTRACT

The purpose of this study was to determine whether it is both possible and cost-effective to revise middle-difficulty Graduate Record Examinations (GRE) discrete items in order to produce items of higher or lower difficulty. The basic procedure was to select items of a given difficulty and, by revising the distractors, make them easier or more difficult. It was found that it is significantly easier to increase the difficulty of middle-difficulty items than to reduce the difficulty of such items and that the difficulties of antonyms and analogies are much easier to manipulate than those of sentence completions. The evidence also suggests that producing harder analogies and antonyms by revising items in this manner would be a cost-effective procedure. Nine appendixes present definitions of GRE content categories, examples of revised items, statistical analysis results for revised items and the tests, pretest yields, and cost figures. Three tables illustrate the discussion. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 385 587

GRE®
RESEARCH

Item Difficulty Adjustment Study: GRE Verbal Discretes

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Richard Adams
John Carson
Kevin Cureton

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 024.025

April 1993

GRE Board Professional Report No. 89-04P
ETS Research Report 92-79

Educational Testing Service, Princeton, New Jersey

BEST COPY AVAILABLE

Item Difficulty Adjustment Study:

GRE Verbal Discretes

Richard Adams
John Carson
Kevin Cureton

GRE Board Report No. 89-04P

April 1993

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

Educational Testing Service, Princeton, N.J. 08541

The Graduate Record Examinations Board and Educational Testing Service are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

EDUCATIONAL TESTING SERVICE, ETS, the ETS logo, GRADUATE RECORD EXAMINATIONS, and GRE are registered trademarks of Educational Testing Service.

Copyright © 1993 by Educational Testing Service. All rights reserved.

Abstract

The purpose of this study was to determine whether it is both possible and cost-effective to revise middle-difficulty GRE discrete items in order to produce items of higher or lower difficulty. It was found that it is significantly easier to increase the difficulty of middle-difficulty items than to reduce the difficulty of such items, and that the difficulties of antonyms and analogies are much easier to manipulate than those of sentence completions. The evidence also suggests that producing harder analogies and antonyms by revising items in this manner would be a cost-effective procedure.

Introduction

Over the past several years, assemblers of the verbal section of the GRE General Test have experienced increasing problems in meeting the current standard deviation of the delta specification, that is, meeting the statistical specification that calls for a relatively wide range of both hard and easy items in a verbal final form. These problems are the result of two somewhat related factors: a continuing inability to obtain, predictably and consistently, discrete items (analogies, antonyms, and sentence completions) that are very easy (Δ 6-8), and a similar, though even more serious, inability to obtain difficult discrete items (Δ 14-17).[1] In addition, the problems of assembly have been further complicated by losses of discrete items caused by the differential item functioning (DIF) statistic, a statistical screening for items that function differentially for subgroups of examinees; these losses, too, seem greatest at the extremes of the delta scale. The situation has, in fact, reached the point that assembly of the October 1989 final forms was delayed in order to wait for high delta items from the October 1988 pretests, because there was an insufficient pool of hard and easy items available for assembly. It is apparent that, even though modification of the standard deviation specification is under study and that statistical specifications based on item response theory (IRT) may relieve the problem to some degree in the future, the need to obtain items of a targeted difficulty, especially high delta items, is now critical and will remain so for the immediate future. The purpose of this study, therefore, is to address the continuing need for items with particular statistical characteristics (difficulty level) in the verbal measure. The study includes attempts at both raising and lowering difficulty levels of middle difficulty items (generating hard and easy items).

Method

A large pool of middle-difficulty, pretested discrete items already exists and, under current pretesting conditions, can never be exhausted in assembling GRE final forms. This pool of items provided the material for the study. The basic procedure was to select items of difficulty Δ 9-10 and, by revising the distractors, make such items easier; similarly, items of difficulty Δ 12-13 were revised in an effort to make them more difficult. Items were selected as follows:

---

[1]Delta is an index of item difficulty based upon the percent of all candidates trying the item who answered it correctly. The principle advantage of delta over P_+ (percent correct) lies in the fact that equal increments in delta, unlike P_+, may reasonably be assumed to represent equal increments in difficulty. This characteristic linearity of delta permits comparisons to be made between groups taking different test forms. A theoretical average delta for a 5-choice item is 12.0. (If 50% of the candidates know the answer and the other 50% answer by chance, 60% of the students will answer the item correctly. A P_+ of .60 corresponds to a delta of 12.0) Deltas ordinarily range from 6.0 for very easy items (approximately 95% correct) to 13.0 for middle difficulty items (approximately 50% correct) to 20.0 for a very hard item (approximately 5% correct). The average range for the GRE verbal measure is 6.5 to 16.5, with a mean for the test of 12.0.

General Content

Type	Content Classification[2]	Delta Range	Number of Items	Intended Outcome
Sentence Completions	Art-Humanities	9-10	5	lower delta
	Art-Humanities	12-13	10	higher delta
	Human Rels	9-10	5	lower delta
	Human Rels	12-13	10	higher delta
Antonyms	Art-Humanities	9-10	5	lower delta
	Art-Humanities	12-13	10	higher delta
	Human Rels	9-10	5	lower delta
	Human Rels	12-13	10	higher delta
Analogies	Art-Humanities	9-10	5	lower delta
	Art-Humanities	12-13	10	higher delta
	Human Rels	9-10	5	lower delta
	Human Rels	12-13	10	higher delta

Item Characteristics

In the case of attempts to reduce the difficulty level of verbal discretes, items were selected that had at least two strong distractors, that is, distractors that were attracting a significant number of test-takers. Distractors were weakened, in general, by relying on the judgments of experienced item writers to lower the vocabulary level and reduce the closeness of the distractor to the credited option. Specifically, for antonyms, words were chosen as distractors that were less semantically appropriate for either the stem or answer contexts; for sentence completion items, words were chosen as distractors that were less likely to appear in the contexts of the given sentences; and for analogy distractors, word pairs were chosen whose analogical relationships were less similar to those of the stem-answer pairs. (See Appendix B for examples of the revisions.)

For increasing the difficulty level of verbal discretes, items were selected that had at least two weak distractors, that is, distractors that were attracting relatively few test takers. Strengthening a distractor, in general, consisted of relying on the judgment of experienced item writers to raise the vocabulary level and increase the closeness of the distractor to the credited option. For antonyms, words were chosen as distractors that were more semantically appropriate for either stem or answer contexts; for sentence completion items, words were chosen as

[2]These classifications have been selected as those that experience shows are least likely to yield high DIF values. See Appendix A for the definitions of the entire set of content categories for GRE verbal discretes.

distractors that were more likely to appear in the contexts of the sentences; and for analogy distractors, pairs of words were chosen whose analogical relationships were more similar to those of the stem-answer pairs. (See Appendix C for examples of the revisions.)

The revised items were assembled into three experimental sections having the following characteristics:

(1) Each section contained a long and short reading comprehension set typical of GRE verbal pretests and final forms; that is, each section maintained the appearance of a standard GRE pretest/final form.

(2) Each section contained both easy and difficult discrete items, although not in the same proportions as in GRE verbal pretests and final forms.

(3) Each section contained several science and practical affairs items in addition to the human relations and arts and humanities items being studied, again to preserve the appearance of the experimental sections as standard verbal pretests.

Results

The complete quantitative results of this study can be found in Appendix D and Appendix E. Table 1 presents a summary of the most significant data, based on an analysis of those items (67 out of 72, or 93%) that had acceptable correlation coefficients ($r_{bis} \geq .20$) after pretesting. All had acceptable r_{bis} before revision.

Our general finding is that manipulating the difficulty of a verbal discrete item in a desired direction by changing some of the distractors is possible. The ability to manipulate the difficulty, however, depends on a number of factors. First, as Table 1 shows, the average reductions in delta achieved during this study (-0.33) were smaller than the average increases in delta that were achieved (+0.78). Second, the difficulty levels of sentence completions (SNCP) proved significantly less manipulable than those of antonyms (ANT) or especially analogies (ANAL). Looking only at those items that were intended to be made harder, it can be seen in Table 1 that, whereas the difficulty levels of sentence completions were raised on average by only +0.28 delta points, the difficulty levels of antonyms and analogies together showed an average increase of +0.97 delta points. In addition, the standard deviations of the deltas of analogies and antonyms indicate that about 66% of all tries yielded increases of between 0 and 2 delta points.

We also compared the cost-effectiveness of the experimental method for producing difficult discretes with the current method of pretesting. Two general methods of assessing the cost/benefits were used. First, we attempted to compare the dollar value ($value) of the items produced by each method with the cost of production for each method. We assumed first that the experimental method would be used only on antonyms and analogies and only to increase deltas. We then used the frequency data given in Appendix F to determine what distribution of deltas, on average, could be expected from a set of 20 analogies and antonyms (one pretest's worth), revised from items with difficulties between Δ12 and Δ13 and having a mean delta of 12.6, and compared that to the standard yields of a normal pretest (the data from which the normal

Table 1
Statistical Results For Acceptable Items (N=67)

Objective	Item Type	Avg Change in Delta	Range	StdDev
Make Easier	SNCP	-0.26	-1.4 to 1.5	0.92
	ANAL	-0.53	-2.1 to 1.1	1.00
	ANT	-0.25	-2.0 to 2.0	1.44
	ANAL+ANT	-0.38		
	AVG	-0.33		
Make Harder	SNCP	+0.28	-2.9 to 1.6	1.39
	ANAL	+1.10	+0.2 to 3.5	1.07
	ANT	+0.88	-1.0 to 3.1	0.99
	ANAL+ANT	+0.97		
	AVG	+0.78		
Control (*N*=26)		+0.12		

Distribution of Items Whose Statistical Change Was in Intended Direction

ΔChange	0.3-0.7	0.8-1.2	1.3-1.7	1.8-2.2	2.3-2.7	2.8-3.2	3.3-3.7	
# of items	8	15	9	5	0	3	1	= 41

and compared that to the standard yields of a normal pretest (the data from which the normal yields have been calculated are contained in Appendix G). Table 2 compares the two sets of yields.

As Table 2 demonstrates, the experimental method produces difficult items at significantly higher rates than those produced by standard pretesting. Using the calculation of the $values of GRE items by delta range found in Appendix H, the $values of the items produced by standard pretesting and by the experimental procedure can be compared (see Table 3). These values show that for one pretest's worth of analogies and antonyms (N=20), the experimental procedure produces items of 42% greater value than those produced by the standard pretesting procedure.

Table 2
Distribution of Deltas for 20 Hard Analogies and Antonyms: Standard versus Experimental

Delta	Standard Frequency (Good R_{bis})	Experimental Frequency (Good R_{bis})
12	2.8	5.7
13	2.2	6.4
14	2.0	2.9
15	1.1	2.1
$\geq$16	0.5	0.7

A comparison of the costs of production for each of these methods, however, yields unclear evidence about the cost-effectiveness of the experimental method. Because the experimental items can be produced more quickly than items produced through standard pretesting (1.2 hours/exp.item vs. 1.8 hours/std.item), test development costs for pretesting 20 analogies and antonyms are $850 less for the experimental method than the costs of the standard method ($7,660 exp. cost vs. $8,550 std. cost). However, before a cost/benefits ratio can be calculated, the value of the pretested items used must be factored into the costs. For standard pretesting, this value is $0 (because there are no pretested items), yielding a cost/benefits ratio of $8,550:$13,457 or $1:$1.57. (See Table 3.)

For the experimental items, initial value can be calculated in a number of ways. Given the large GRE pool of Δ12 analogies and antonyms and the fact that in a standard year of 20 pretests and 4 final forms 100% more Δ12s and 24% more Δ13s will be produced than will be needed, one reasonable assessment of the initial value of the discretes used in the experimental method is also $0. This results in a cost/benefits ratio of $7,660:$19,173 or $1:$2.50. Using the determination of values of GRE discretes of various deltas listed in Appendix H, however, results in an initial value for the items used in the experimental method of $10,985 (14 Δ12s, and 6 Δ13s), and a cost/benefits ratio of $18,645:$19,173 or $1:$1.03. Thus, using the first estimation of initial value, the experimental method produces approximately one-half times more benefit per dollar spent than does the standard method. By the second estimation of initial value, however, the experimental method produces only about the same benefit per dollar spent as does the standard method. We believe that estimating the initial value of the revised items at $0 is the more reasonable way to proceed, because large portions of the GRE pool of middle-difficulty items will, in all probability, never be used.

A second way to calculate the cost/benefits of the experimental method is to determine how many fewer pretests could be run if a mixture of standard method pretests and experimental method pretests were used. Using the data found in Appendix I, it was determined that a mixture of 14 standard pretests and 3 experimental pretests would best yield the requisite number of items

Table 3
Value Produced in Pretesting 20 Analogies and Antonyms:
Standard versus Experimental Method
(Items with Good R_{bis})

Delta	$Value/ Delta*	Std Freq**	Exp Freq***	Std $Value	Exp $Value
<8	757	0.6	0.0	482	0
8	757	1.1	0.0	826	0
9	892	1.8	0.0	1,621	0
10	245	2.0	0.0	491	0
11	203	3.3	0.7	665	145
12	403	2.8	5.7	1,134	2,300
13	892	2.2	6.4	1,945	5,731
14	757	2.0	2.9	1,515	2,164
15	3,092	1.1	2.1	3,373	6,625
≥16	3,092	0.5	0.7	1,405	2,208
TOTAL				$13,457	$19,173

*See Appendix H. **See Appendix G. ***See Appendix F.

at the various deltas where they are needed.[3] If the value of the pretested items used in the experimental pretests is set at $0, the savings per year would be $62,550 ($20,000 [T.D. costs/pretest] x 3 [pretests saved/yr] + $850 [T.D. savings/exp.pretest] x 3 [exp. pretests/yr]). If the values listed in Appendix H are used, the savings per year would be $29,596 ($62,550 - $10,985 [value of items/exp. pretest] x 3 [exp. pretests/yr]).

The preceding shows that a consistent estimate of the savings that could be achieved by including the experimental method in GRE pretesting is difficult to come by, mainly because there is no one obvious way to determine the value of the already pretested items reused in the experimental pretests. Our best guess, using the figures for pretest savings, is that a mixture of standard and experimental pretests could save the GRE programs between $30,000 and $60,000 per year if we can extrapolate from the data on which our study is based, and if we can safely cut

[3]Because the experimental items were written in content categories chosen to minimize DIF, it is possible that the yields listed in Appendix I for the experimental pretests might in practice be lower. However, because the pool of pretested items from which the experimental items were picked was deliberately restricted to older items for which no DIF data were available, it is likely that these experimental items show more DIF than would occur if revision was made of items for which DIF information was already known.

back to 17 pretests per year.

Conclusions

The results of our research indicate that the experimental procedure can be used to enrich the GRE discrete pool, but only with the following provisos: (1) The method should be used only to produce higher delta items; and (2) the method should not be applied to sentence completion items. We believe that this method is likely to be a cost-effective way to produce more difficult analogy and antonym discretes, and that it can be effectively introduced by dedicating three pretests worth of analogies and antonyms per year to items revised in the ways suggested in this study.

We also believe that further research is in order before the benefits of this method are certain. We recommend (1) trying a more extensive study using all four content categories and focusing specifically on increasing the difficulty of antonyms and analogies; (2) studying the ability to increase the difficulty of discretes that are not middle range in difficulty (i.e., is it possible by using the same techniques to raise Δ14 and Δ15 items to Δ15, Δ16, and Δ17 items?); (3) doing a more extensive investigation of the possibility of making Δ10 and Δ11 analogies easier using this method; and (4) using this method of research to determine, for each item type, the degree to which distractors affect item performance (i.e., it seems from data on SNCP items that performance on those items is not much affected by the distractors).

APPENDIX A: DEFINITIONS OF GRE CONTENT CATEGORIES

Art-Humanities

This category includes words (or, in the case of sentence completions, issues and ideas) that are most likely to be used or discussed in the discourse of, or discourse about, the fine and applied arts (painting and architecture, for example), literature, philosophy, religion, and other such fields.

Social Studies and Practical or Everyday Life

This category includes words (or, in the cases of sentence completions, issues and ideas) that are most likely to be used or discussed in the discourse of, or discourse about, such fields as communications, business, politics and government, economics, and transportation.

Science and Nature

This category includes words (or, in the case of sentence completions, issues and ideas) that are most likely to be used or discussed in the discourse of, or discourse about, such fields as biology, chemistry, physics, geology, and the theoretical and applied sciences, such as mathematics and medicine.

Human Relationships and Feelings

This category includes words (or, in the case of sentence completions, issues and ideas) that concern emotions, interpersonal relationships, or analyses of character.

APPENDIX B: EXAMPLES OF ITEMS REVISED TO BECOME EASIER

ITEM: V-081807

Original Item

STANZA:POEM::

(A) play:drama
(B) lyric:song
(C) chapter:book
(D) stone:statue
(E) reproduction:painting

BASE N	OMIT	A	B	C *	D	E	M-TOT	SCALE	ΔE	CRIT
1,995	91	21	349	1511	16	5	13.0	NGR1	9.4	XS80
ITEM #	**M-O**	**M-A**	**M-B**	**M-C**	**M-D**	**M-E**	**P-TOT**	**P+**	**ΔO**	**Rbis**
14	8.3	6.5	11.2	13.9	8.6	8.0	1.00	0.76	10.2	0.52

Revised Item

STANZA:POEM::

(A) play:drama
(B) music:song
(C) chapter:book
(D) stone:statue
(E) reproduction:painting

BASE N	OMIT	A	B	C *	D	E	M-TOT	SCALE	ΔE	CRIT
1,045	0	10	18	1000	15	2	13.0	3DGR	7.3	XS76
ITEM #	**M-O**	**M-A**	**M-B**	**M-C**	**M-D**	**M-E**	**P-TOT**	**P+**	**ΔO**	**Rbis**
8	0.0	6.5	7.7	13.2	9.7	11.0	1.00	0.96		0.53

* indicates key

ITEM: V-096399

Original Item

RESENT:

(A) accommodate
(B) welcome
(C) protect
(D) concern
(E) insist

BASE N	OMIT	A	B *	C	D	E	M-TOT	SCALE	ΔE	CRIT
1,470	4	154	1098	55	80	42	13.1	3DGR	9.8	XS76
ITEM #	**M-O**	**M-A**	**M-B**	**M-C**	**M-D**	**M-E**	**P-TOT**	**P+**	**ΔO**	**Rbis**
30	10.0	12.1	13.6	10.7	11.0	10.2	0.97	0.77	10.1	0.33

Revised Item

RESENT:

(A) hasten
(B) welcome
(C) protect
(D) employ
(E) demand

BASE N	OMIT	A	B *	C	D	E	M-TOT	SCALE	ΔE	CRIT
1,045	1	10	970	12	42	9	13.0	3DGR	8.1	XS76
ITEM #	**M-O**	**M-A**	**M-B**	**M-C**	**M-D**	**M-E**	**P-TOT**	**P+**	**ΔO**	**Rbis**
28	10.0	7.8	13.2	10.3	10.4	9.9	1.00	0.93	7.1	0.40

ITEM: V-081813

Original Item

Even though most dance critics believe that the choreographer succeeded in ------- the public's notions of how dance should be performed, her radical ideas have in fact not yet been fully -------.

(A) contradicting..revealed
(B) challenging..negated
(C) reshaping..accepted
(D) capturing..developed
(E) confirming..understood

BASE N	OMIT	A	B	C *	D	E	M-TOT	SCALE	ΔE	CRIT
1,995	40	44	41	1468	227	173	13.0	NGR1	9.7	XS80
ITEM #	M-O	M-A	M-B	M-C	M-D	M-E	P-TOT	P+	ΔO	Rbis
19	8.5	11.4	9.9	13.8	10.9	11.3	1.00	0.74	10.4	0.45

Revised Item

Even though most dance critics believe that the choreographer succeeded in ------- the public's notions of how dance should be performed, her radical ideas have in fact not yet been fully -------.

(A) presenting..arranged
(B) challenging..negated
(C) reshaping..accepted
(D) controlling..exercised
(E) underestimating..regulated

BASE N	OMIT	A	B	C *	D	E	M-TOT	SCALE	ΔE	CRIT
1,045	1	32	30	962	18	2	13.0	3DGR	8.3	XS76
ITEM #	M-O	M-A	M-B	M-C	M-D	M-E	P-TOT	P+	ΔO	Rbis
1	12.0	10.4	9.6	13.3	8.7	8.5	1.00	0.92	7.4	0.44

APPENDIX C: EXAMPLES OF ITEMS REVISED TO BECOME HARDER

ITEM: V-083094

Original Item

LOQUACIOUS:CHATTER::

(A) perilous:safeguard
(B) numerous:count
(C) marvelous:admire
(D) officious:meddle
(E) precious:cherish

BASE N	OMIT	A	B	C	D *	E	M-TOT	SCALE	ΔE	CRIT
1,515	138	115	235	183	590	231	13.1	NGR1	12.6	XS75
ITEM #	M-O	M-A	M-B	M-C	M-D	M-E	P-TOT	P+	ΔO	Rbis
47	10.7	11.8	12.4	11.1	15.0	12.4	0.98	0.40	14.1	0.50

Revised Item

LOQUACIOUS:CHATTER::

(A) meticulous:complain
(B) numerous:count
(C) voracious:starve
(D) officious:meddle
(E) precious:cherish

BASE N	OMIT	A	B	C	D *	E	M-TOT	SCALE	ΔE	CRIT
1,015	3	116	107	224	446	119	13.0	3DGR	13.6	XS76
ITEM #	M-O	M-A	M-B	M-C	M-D	M-E	P-TOT	P+	ΔO	Rbis
16	9.3	10.6	11.6	13.0	14.3	11.8	1.00	0.44	13.6	0.37

ITEM: V-079231

Original Item

DUBITABLE:

(A) essential
(B) certain
(C) sensible
(D) decisive
(E) verifiable

BASE N	OMIT	A	B *	C	D	E	M-TOT	SCALE	ΔE	CRIT
795	262	20	244	29	44	101	13.3	NGR1	12.8	XS80
ITEM #	**M-O**	**M-A**	**M-B**	**M-C**	**M-D**	**M-E**	**P-TOT**	**P+**	**ΔO**	**Rbis**
50	11.7	11.8	15.1	10.1	11.5	14.9	0.88	0.35	14.8	0.44

Revised Item

DUBITABLE:

(A) reassuring
(B) verifiable
(C) coercive
(D) decisive
(E) certain

BASE N	OMIT	A	B	C	D	E *	M-TOT	SCALE	ΔE	CRIT
1,015	6	127	250	150	99	357	13.1	3DGR	14.3	XS76
ITEM #	**M-O**	**M-A**	**M-B**	**M-C**	**M-D**	**M-E**	**P-TOT**	**P+**	**ΔO**	**Rbis**
36	12.7	12.4	13.3	10.3	11.0	14.8	0.97	0.36	14.4	0.43

ITEM: V-099427

Original Item

In the midst of so many evasive comments, this forthright statement, whatever its intrinsic merit, plainly stands out as -------.

(A) an anomaly
(B) an inaccuracy
(C) an inference
(D) a misnomer
(E) a pattern

BASE N	OMIT	A *	B	C	D	E	M-TOT	SCALE	ΔE	CRIT
1,100	3	628	110	140	176	43	13.0	3DGR	11.9	XS76
ITEM #	**M-O**	**M-A**	**M-B**	**M-C**	**M-D**	**M-E**	**P-TOT**	**P+**	**ΔO**	**Rbis**
3	8.3	14.8	10.6	10.0	11.6	8.3	1.00	0.57	12.3	0.67

Revised Item

In the midst of so many evasive comments, this forthright statement, whatever its intrinsic merit, plainly stands out as -------.

(A) an anomaly
(B) an inaccuracy
(C) a profundity
(D) a misnomer
(E) a paradigm

BASE N	OMIT	A *	B	C	D	E	M-TOT	SCALE	ΔE	CRIT
1,045	2	471	78	254	90	150	13.0	3DGR	13.5	XS76
ITEM #	**M-O**	**M-A**	**M-B**	**M-C**	**M-D**	**M-E**	**P-TOT**	**P+**	**ΔO**	**Rbis**
6	11.0	14.8	10.1	12.4	10.3	11.5	1.00	0.45	13.5	0.52

APPENDIX D: STATISTICAL RESULTS FOR ALL ITEMS (*N*=72)

Objective	Item Type	Avg Change in Delta	Range	Std Dev
MakeEASIER	SNCP	-0.26	-2.1 to 1.1	0.92
	ANAL	-0.49	-1.5 to 1.1	0.96
	ANT	-0.16	-2.0 to 2.0	1.41
	AVG	-0.33		
MakeHARDER	SNCP	+0.30	-2.9 to 2.8	1.32
	ANAL	+1.29	+0.2 to 3.5	1.20
	ANT	+0.95	-1.0 to 3.1	0.99
	AVG	+0.78		
CONTROL	**AVG**	+0.12		

Objective	Form	Avg Change in Delta
MakeEASIER	X1	+0.37
	X2	-0.29
	X3	-1.08
MakeHARDER	X1	+0.35
	X2	+1.38
	X3	+0.82

DATA ON ALL EXPERIMENTAL ITEMS

	ACC#	FORM	ITEM#	TYPE	HARD/EASY	ORIG DELTA	NEW DELTA	CHG	NEW RBIS
	V-099004	X1	2	SNCP	Easier	9.0	10.5	1.5	0.34
	V-076475	X1	3	SNCP	Easier	9.4	8.9	-0.5	0.52
	V-083010	X1	4	SNCP	Easier	9.6	10.7	1.1	0.42
	V-076482	X2	1	SNCP	Easier	9.0	8.1	-0.9	0.45
*	V-074935	X2	2	SNCP	Easier	10.6	10.1	-0.5	0.53
	V-085023	X2	3	SNCP	Easier	10.8	9.9	-0.9	0.21
	V-081813	X3	1	SNCP	Easier	9.7	8.3	-1.4	0.44
	V-053534	X3	2	SNCP	Easier	9.8	9.9	0.1	0.59
	V-065681	X3	3	SNCP	Easier	10.7	9.9	-0.8	0.52
	V-078014	X1	8	ANAL	Easier	9.1	10.1	1.0	0.37
	V-083088	X1	9	ANAL	Easier	9.4	8.5	-0.9	0.61
	V-089162	X1	10	ANAL	Easier	9.6	8.3	-1.3	0.39
	V-066778	X1	12	ANAL	Easier	10.7	11.8	1.1	0.37
*	V-081746	X2	8	ANAL	Easier	10.0	8.5	-1.5	0.21
	V-069672	X2	9	ANAL	Easier	10.6	9.6	-1.0	0.53
	V-099544	X2	10	ANAL	Easier	10.6	10.5	-0.1	0.17
	V-087748	X2	12	ANAL	Easier	10.9	11.0	0.1	0.42
	V-081807	X3	8	ANAL	Easier	9.4	7.3	-2.1	0.53
*	V-096379	X3	9	ANAL	Easier	10.6	10.6	0.0	0.53
	V-096439	X3	10	ANAL	Easier	10.7	10.0	-0.7	0.42
*	V-083067	X1	28	ANT	Easier	9.5	10.4	0.9	0.07
*	V-083020	X1	30	ANT	Easier	10.1	11.8	1.7	0.57
	V-095530	X1	31	ANT	Easier	10.0	10.4	0.4	0.55
	V-089430	X1	32	ANT	Easier	10.5	9.3	-1.2	0.58
	V-081310	X2	28	ANT	Easier	9.4	11.4	2.0	0.25
*	V-097401	X2	29	ANT	Easier	10.0	11.5	1.5	0.49
	V-056614	X2	30	ANT	Easier	10.1	10.7	0.6	0.46
	V-076500	X2	32	ANT	Easier	10.5	10.8	0.3	0.64
	V-087868	X2	33	ANT	Easier	10.7	8.9	-1.8	0.70
	V-096399	X3	28	ANT	Easier	9.8	8.1	-1.7	0.40
	V-079173	X3	29	ANT	Easier	10.1	8.1	-2.0	0.51
	V-083572	X3	30	ANT	Easier	10.0	8.1	-1.9	0.29
*	V-087499	X3	31	ANT	Easier	10.9	10.1	-0.8	0.48

*DIF B item **DIF C item

	ACC#	FORM	ITEM#	TYPE	HARD/EASY	ORIG DELTA	NEW DELTA	CHG	NEW RBIS
	V-074937	X1	5	SNCP	Harder	11.0	11.2	0.2	0.45
	V-095557	X1	6	SNCP	Harder	11.0	10.8	-0.2	0.51
	V-086165	X1	7	SNCP	Harder	11.2	11.4	0.2	0.44
	V-087902	X2	4	SNCP	Harder	11.1	11.3	0.2	0.54
	V-099392	X2	5	SNCP	Harder	11.2	8.3	-2.9	0.39
	V-082513	X2	6	SNCP	Harder	11.3	14.1	2.8	0.51
	V-067413	X2	7	SNCP	Harder	12.1	12.6	0.5	0.12
	V-077947	X3	4	SNCP	Harder	11.0	11.8	0.8	0.35
	V-076638	X3	5	SNCP	Harder	11.1	11.6	0.5	0.49
	V-099427	X3	6	SNCP	Harder	11.9	13.5	1.6	0.52
	V-087504	X3	7	SNCP	Harder	13.1	12.7	-0.4	0.29
	V-058701	X1	13	ANAL	Harder	12.1	12.3	0.2	0.26
*	V-099419	X1	14	ANAL	Harder	12.3	12.8	0.5	0.55
	V-086169	X1	15	ANAL	Harder	12.4	12.4	0.0	0.21
	V-083094	X1	16	ANAL	Harder	12.6	13.6	1.0	0.37
	V-092972	X2	13	ANAL	Harder	12.1	13.0	0.9	0.29
**	V-065697	X2	14	ANAL	Harder	12.2	13.6	1.4	0.30
	V-078012	X2	15	ANAL	Harder	13.1	16.5	3.4	0.07
	V-093062	X2	16	ANAL	Harder	13.4	16.3	2.9	0.20
	V-099492	X3	13	ANAL	Harder	12.0	12.9	0.9	0.35
	V-083650	X3	14	ANAL	Harder	12.1	12.6	0.5	0.50
	V-079197	X3	15	ANAL	Harder	12.3	15.8	3.5	0.29
	V-096453	X3	16	ANAL	Harder	13.1	13.4	0.3	0.33
*	V-099011	X1	34	ANT	Harder	12.0	12.9	0.9	0.62
	V-083721	X1	35	ANT	Harder	12.1	12.3	0.2	0.47
	V-079231	X1	36	ANT	Harder	12.8	14.3	1.5	0.43
**	V-093081	X1	37	ANT	Harder	13.3	13.4	0.1	0.52
	V-092619	X1	38	ANT	Harder	13.4	13.5	0.1	0.36
	V-094672	X2	34	ANT	Harder	12.0	15.1	3.1	0.21
	V-092815	X2	35	ANT	Harder	12.1	13.4	1.3	0.23
	V-097403	X2	36	ANT	Harder	12.8	15.0	2.2	0.23
	V-083069	X2	37	ANT	Harder	12.9	14.5	1.6	0.30
	V-087508	X2	38	ANT	Harder	13.1	14.1	1.0	0.55
**	V-098409	X3	33	ANT	Harder	12.2	11.8	-0.4	0.48
	V-087871	X3	34	ANT	Harder	12.3	13.0	0.7	0.47
	V-042924	X3	35	ANT	Harder	13.0	13.9	0.9	0.35
	V-096313	X3	36	ANT	Harder	13.3	14.3	1.0	0.43
	V-092606	X3	37	ANT	Harder	13.4	15.4	2.0	-0.04
	V-098334	X3	38	ANT	Harder	13.6	12.6	-1.0	0.20

*DIF B item **DIF C item

ACC#	FORM	ITEM#	TYPE	HARD/EASY	ORIG DELTA	NEW DELTA	CHG	NEW RBIS
V-098371	X1	11	ANAL	Control	9.6	9.6	0.0	0.41
V-074995	X2	11	ANAL	Control	10.6	11.0	0.4	0.43
V-099543	X3	12	ANAL	Control	11.8	12.3	0.5	0.37
V-075040	X1	33	ANT	Control	10.6	10.9	0.3	0.58
V-099563	X1	17	RCMP	Control	11.0	10.2	-0.8	0.60
V-099564	X1	18	RCMP	Control	13.5	13.5	0.0	0.39
V-099565	X1	19	RCMP	Control	13.6	13.5	-0.1	0.48
V-099566	X1	20	RCMP	Control	14.0	13.0	-1.0	0.40
V-087794	X1	21	RCMP	Control	10.6	10.4	-0.2	0.35
V-087796	X1	22	RCMP	Control	9.5	9.7	0.2	0.49
V-087800	X1	23	RCMP	Control	12.3	12.1	-0.2	0.44
V-087799	X1	24	RCMP	Control	11.3	10.7	-0.6	0.58
V-087802	X1	25	RCMP	Control	13.5	14.1	0.6	0.45
V-087795	X1	26	RCMP	Control	11.6	11.8	0.2	0.58
BE000616	X1	27	RCMP	Control	11.4	11.2	-0.2	0.38
IF000114	X2	18	RCMP	Control	14.2	14.3	0.1	0.37
IF000115	X2	19	RCMP	Control	13.2	13.0	-0.2	0.45
IF000119	X2	21	RCMP	Control	14.3	14.9	0.6	0.30
IF000133	X2	22	RCMP	Control	13.7	13.7	0.0	0.24
V-093692	X2	25	RCMP	Control	14.3	13.8	-0.5	0.43
V-093693	X2	26	RCMP	Control	14.4	14.2	-0.2	0.33
V-093694	X2	27	RCMP	Control	12.8	13.1	0.3	0.39
HT000415	X3	24	RCMP	Control	8.1	9.1	1.0	0.47
HT000417	X3	25	RCMP	Control	11.7	12.7	1.0	0.44
HT000438	X3	26	RCMP	Control	8.2	9.3	1.1	0.41
HT000418	X3	27	RCMP	Control	10.1	10.9	0.8	0.15

*DIF B item **DIF C item

APPENDIX E: STATISTICAL RESULTS FOR ACCEPTABLE (RBIS≥.20) ITEMS (*N*=67)

Objective	Item Type	Avg Change in Delta	Range	Std Dev
MakeEASIER	SNCP	-0.26	-1.4 to 1.5	0.92
	ANAL	-0.53	-2.1 to 1.1	1.00
	ANT	-0.25	-2.0 to 2.0	1.44
	ANAL+ANT	-0.38		
	AVG	-0.33		
MakeHARDER	SNCP	+0.28	-2.9 to 1.6	1.39
	ANAL	+1.10	+0.2 to 3.5	1.07
	ANT	+0.88	-1.0 to 3.1	0.99
	ANAL+ANT	+0.97		
	AVG	+0.78		
CONTROL	**AVG**	+0.12		

Distribution of Intended Changes in Delta

ΔChange	**0.3-0.7**	**0.8-1.2**	**1.3-1.7**	**1.8-2.2**	**2.3-2.7**	**2.8-3.2**	**3.3-3.7**
# of items	8	15	9	5	0	3	1

DATA ON ACCEPTABLE EXPERIMENTAL ITEMS

	ACC#	FORM	ITEM#	TYPE	HARD/EASY	ORIG DELTA	NEW DELTA	CHG	NEW RBIS
	V-099004	X1	2	SNCP	Easier	9.0	10.5	1.5	0.34
	V-076475	X1	3	SNCP	Easier	9.4	8.9	-0.5	0.52
	V-083010	X1	4	SNCP	Easier	9.6	10.7	1.1	0.42
	V-076482	X2	1	SNCP	Easier	9.0	8.1	-0.9	0.45
*	V-074935	X2	2	SNCP	Easier	10.6	10.1	-0.5	0.53
	V-085023	X2	3	SNCP	Easier	10.8	9.9	-0.9	0.21
	V-081813	X3	1	SNCP	Easier	9.7	8.3	-1.4	0.44
	V-053534	X3	2	SNCP	Easier	9.8	9.9	0.1	0.59
	V-065681	X3	3	SNCP	Easier	10.7	9.9	-0.8	0.52
	V-078014	X1	8	ANAL	Easier	9.1	10.1	1.0	0.37
	V-083088	X1	9	ANAL	Easier	9.4	8.5	-0.9	0.61
	V-089162	X1	10	ANAL	Easier	9.6	8.3	-1.3	0.39
	V-066778	X1	12	ANAL	Easier	10.7	11.8	1.1	0.37
*	V-081746	X2	8	ANAL	Easier	10.0	8.5	-1.5	0.21
	V-069672	X2	9	ANAL	Easier	10.6	9.6	-1.0	0.53
	V-087748	X2	12	ANAL	Easier	10.9	11.0	0.1	0.42
	V-081807	X3	8	ANAL	Easier	9.4	7.3	-2.1	0.53
*	V-096379	X3	9	ANAL	Easier	10.6	10.6	0.0	0.53
	V-096439	X3	10	ANAL	Easier	10.7	10.0	-0.7	0.42
*	V-083020	X1	30	ANT	Easier	10.1	11.8	1.7	0.57
	V-095530	X1	31	ANT	Easier	10.0	10.4	0.4	0.55
	V-089430	X1	32	ANT	Easier	10.5	9.3	-1.2	0.58
	V-081310	X2	28	ANT	Easier	9.4	11.4	2.0	0.25
*	V-097401	X2	29	ANT	Easier	10.0	11.5	1.5	0.49
	V-056614	X2	30	ANT	Easier	10.1	10.7	0.6	0.46
	V-076500	X2	32	ANT	Easier	10.5	10.8	0.3	0.64
	V-087868	X2	33	ANT	Easier	10.7	8.9	-1.8	0.70
	V-096399	X3	28	ANT	Easier	9.8	8.1	-1.7	0.40
	V-079173	X3	29	ANT	Easier	10.1	8.1	-2.0	0.51
	V-083572	X3	30	ANT	Easier	10.0	8.1	-1.9	0.29
*	V-087499	X3	31	ANT	Easier	10.9	10.1	-0.8	0.48

*DIF B item **DIF C item

	ACC#	FORM	ITEM#	TYPE	HARD/EASY	ORIG DELTA	NEW DELTA	CHG	NEW RBIS
	V-074937	X1	5	SNCP	Harder	11.0	11.2	0.2	0.45
	V-095557	X1	6	SNCP	Harder	11.0	10.8	-0.2	0.51
	V-086165	X1	7	SNCP	Harder	11.2	11.4	0.2	0.44
	V-087902	X2	4	SNCP	Harder	11.1	11.3	0.2	0.54
	V-099392	X2	5	SNCP	Harder	11.2	8.3	-2.9	0.39
	V-082513	X2	6	SNCP	Harder	11.3	14.1	2.8	0.51
	V-077947	X3	4	SNCP	Harder	11.0	11.8	0.8	0.35
	V-076638	X3	5	SNCP	Harder	11.1	11.6	0.5	0.49
	V-099427	X3	6	SNCP	Harder	11.9	13.5	1.6	0.52
	V-087504	X3	7	SNCP	Harder	13.1	12.7	-0.4	0.29
	V-058701	X1	13	ANAL	Harder	12.1	12.3	0.2	0.26
*	V-099419	X1	14	ANAL	Harder	12.3	12.8	0.5	0.55
	V-086169	X1	15	ANAL	Harder	12.4	12.4	0.0	0.21
	V-083094	X1	16	ANAL	Harder	12.6	13.6	1.0	0.37
	V-092972	X2	13	ANAL	Harder	12.1	13.0	0.9	0.29
**	V-065697	X2	14	ANAL	Harder	12.2	13.6	1.4	0.30
	V-093062	X2	16	ANAL	Harder	13.4	16.3	2.9	0.20
	V-099492	X3	13	ANAL	Harder	12.0	12.9	0.9	0.35
	V-083650	X3	14	ANAL	Harder	12.1	12.6	0.5	0.50
	V-079197	X3	15	ANAL	Harder	12.3	15.8	3.5	0.29
	V-096453	X3	16	ANAL	Harder	13.1	13.4	0.3	0.33
*	V-099011	X1	34	ANT	Harder	12.0	12.9	0.9	0.62
	V-083721	X1	35	ANT	Harder	12.1	12.3	0.2	0.47
	V-079231	X1	36	ANT	Harder	12.8	14.3	1.5	0.43
**	V-093081	X1	37	ANT	Harder	13.3	13.4	0.1	0.52
	V-092619	X1	38	ANT	Harder	13.4	13.5	0.1	0.36
	V-094672	X2	34	ANT	Harder	12.0	15.1	3.1	0.21
	V-092815	X2	35	ANT	Harder	12.1	13.4	1.3	0.23
	V-097403	X2	36	ANT	Harder	12.8	15.0	2.2	0.23
	V-083069	X2	37	ANT	Harder	12.9	14.5	1.6	0.30
	V-087508	X2	38	ANT	Harder	13.1	14.1	1.0	0.55
**	V-098409	X3	33	ANT	Harder	12.2	11.8	-0.4	0.48
	V-087871	X3	34	ANT	Harder	12.3	13.0	0.7	0.47
	V-042924	X3	35	ANT	Harder	13.0	13.9	0.9	0.35
	V-096313	X3	36	ANT	Harder	13.3	14.3	1.0	0.43
	V-098334	X3	38	ANT	Harder	13.6	12.6	-1.0	0.20

*DIF B item **DIF C item

APPENDIX F: FREQUENCY DISTRIBUTION FOR HARD EXPERIMENTAL ANALOGIES AND ANTONYMS WITH ACCEPTABLE RBIS

Delta Range	# Produced in Exp.Pretsts	Freq/ Pretst*	# Produced with Good DIF	Freq/ Pretst*
<11.0	0	0.0	0	0.0
11.0 - 11.9	1	0.7	0	0.0
12.0 - 12.9	8	5.7	6	4.3
13.0 - 13.9	9	6.4	7	5.0
14.0 - 14.9	4	2.9	4	2.9
15.0 - 15.9	3	2.1	3	2.1
≥16.0	1	0.7	1	0.7
Unacceptable Rbis	2	1.4	2	1.4
TOTAL	28		23	

*Frequency/pretest calculated by dividing the number of items produced in each delta range by 28/20 ([total # hard ANALS & ANTS in exp. pretests]/[# ANALS & ANTS per 1 pretest]). This gives the frequency to be expected were an entire pretest's worth of analogies and antonyms revised to be made harder according to the experimental procedure.

APPENDIX G: CUMULATIVE GRE PRETEST STATISTICS
(K-3KGR3$_{M4AAB}$, K-3KGR3$_{M39A-43B}$, and K-3KGR1$_{44A-49B}$)

Tally #1

Number of items with $r_{bis} \geq .25$, by item type and delta

	<8	8	9	10	11	12	13	14	15	≥16	TOT
RCMP	2	8	15	30	46	32	21	10	1	0	165
SNCP	10	10	14	13	9	5	5	7	1	1	75
ANAL	4	3	11	9	19	11	10	5	8	1	81
ANTM	3	9	9	13	17	20	14	17	4	4	110
TOTAL	19	30	49	65	91	68	50	39	14	6	431
YIELD	1.7	2.7	4.5	5.9	8.3	6.2	4.5	3.5	1.3	0.5	38.9

Tally #2

Number of items with $r_{bis} \geq .25$ and acceptable DIF, by item type and delta

	<8	8	9	10	11	12	13	14	15	≥16	TOT
RCMP	2	8	10	28	38	30	13	9	1	0	140
SNCP	9	6	11	9	7	3	4	7	1	1	60
ANAL	3	2	3	5	12	5	8	5	8	1	52
ANTM	2	4	4	8	8	17	9	16	4	4	80
TOTAL	16	20	28	50	65	55	34	37	14	6	332
YIELD	1.5	1.8	2.5	4.5	5.9	5.0	3.1	3.4	1.3	0.5	29.5

Tally #3

Number of items with $r_{bis} < .25$, by item type

RCMP: 15 **SNCP:** 2 **ANAL:** 10 **ANTM:** 11 **TOTAL:** 38

FREQUENCY DISTRIBUTION FOR STANDARD ANALOGIES AND ANTONYMS

Delta Range	# Produced in Pretsts	Freq/ Pretst*	# Produced with Good DIF	Freq/ Pretst*
<8.0	7	0.6	5	0.5
8.0 - 8.9	12	1.1	6	0.5
9.0 - 9.9	20	1.8	7	0.6
10.0 - 10.9	22	2.0	13	1.2
11.0 - 11.9	36	3.3	20	1.8
12.0 - 12.9	31	2.8	22	2.0
13.0 - 13.9	24	2.2	17	1.5
14.0 - 14.9	22	2.0	21	1.9
15.0 - 15.9	12	1.1	12	1.1
≥16.0	5	0.5	5	0.5

*Frequency/pretest calculated by dividing the number of items produced in each delta range by 11 [# of pretests' worth of data the raw figures are based on].

APPENDIX H: VALUES FOR GRE PRETEST ITEMS BY DELTA (MARGINAL UTILITY ANALYSIS)

Delta:	<8	8	9	10	11	12	13	14	15	≥16
Avg # used in 4 Final Forms:	20	24	36	28	32	44	44	44	23	9
Avg yield/ Std.Pretest: (All item types)*	1.5	1.8	2.5	4.5	5.9	5.0	3.1	3.4	1.3	0.5

Pretest Yields	<8	8	9	10	11	12	13	14	15	≥16	Item Used Yield**	Cost/ Item Used***
1 Pretest	1	2	3	5	6	5	3	3	1	1	30	677
2 Pretests	3	4	5	9	12	10	6	7	3	1	30	677
3 Pretests	4	5	8	14	18	15	9	10	4	2	30	677
4 Pretests	6	7	10	18	24	20	12	13	5	2	30	677
5 Pretests	7	9	13	23	30	25	15	17	6	3	30	677
6 Pretests	9	11	15	27	35	30	19	20	8	3	30	677
7 Pretests	10	13	18	32		35	22	24	9	4	24	846
8 Pretests	12	15	20			40	25	27	10	4	19	1,048
9 Pretests	13	16	23			45	28	30	11	5	19	1,048
10 Pretests	15	18	25			50	31	34	13	5	19	1,048
11 Pretests	16	20	28				34	37	14	6	14	1,419
12 Pretests	17	22	31				37	40	15	7	14	1,419
13 Pretests	19	24	33				40	44	17	7	14	1,419
14 Pretests	20	25	36				43	47	18	8	14	1,419
15 Pretests	22	27	38				46	50	19	8	14	1,419
16 Pretests			41				49		20	9	7	2,683
17 Pretests									22	9	2	11,000
18 Pretests									23	10	2	11,000
19 Pretests									24	10	2	11,000
20 Pretests									25	11	2	11,000
TOT $/Δ:****	22,033	27,541	45,386	22,308	24,000	40,253	55,112	50,950	78,693	33,726		
20 Pretests Yield:	29	36	51	91	118	100	62	67	25	11		
AVG $/ΔItem:	757	757	892	245	203	403	892	757	3,092	3,092		

* Data from Appendix G, Tally #2
** # of items in n'th pretest needed to meet target goals for four final forms
*** Calculated by dividing $20,000 [cost/pretest] by (# of items used in n'th pretest)
**** Calculated by summing for each Δ the # of items used in the n'th pretest times the cost/item used

APPENDIX I: STANDARD AND EXPERIMENTAL PRETEST YIELDS (ACCEPTABLE DIF)

STANDARD PRETESTING YIELDS

Delta:	<8	8	9	10	11	12	13	14	15	≥16
Avg # used in 4 Final Forms:	20	24	36	28	32	44	44	44	23	9
Avg yield/ Std.Pretest: (All Item Types)*	1.5	1.8	2.5	4.5	5.9	5.0	3.1	3.4	1.3	0.5

Standard Pretest Yields	<8	8	9	10	11	12	13	14	15	≥16
1 Pretest	1	2	3	5	6	5	3	3	1	1
2 Pretests	3	4	5	9	12	10	6	7	3	1
3 Pretests	4	5	8	14	18	15	9	10	4	2
4 Pretests	6	7	10	18	24	20	12	13	5	2
5 Pretests	7	9	13	23	30	25	15	17	6	3
6 Pretests	9	11	15	27	35	30	19	20	8	3
7 Pretests	10	13	18	32	41	35	22	24	9	4
8 Pretests	12	15	20	36	47	40	25	27	10	4
9 Pretests	13	16	23	41	53	45	28	30	11	5
10 Pretests	15	18	25	45	59	50	31	34	13	5
11 Pretests	16	20	28	50	65	55	34	37	14	6
12 Pretests	17	22	31	55	71	60	37	40	15	7
13 Pretests	19	24	33	59	77	65	40	44	17	7
14 Pretests	20	25	36	64	83	70	43	47	18	8
15 Pretests	22	27	38	68	89	75	46	50	19	8
16 Pretests	23	29	41	73	95	80	49	54	20	9
17 Pretests	25	31	43	77	100	85	53	57	22	9
18 Pretests	26	33	46	82	106	90	56	61	23	10
19 Pretests	28	35	48	86	112	95	59	64	24	10
20 Pretests	29	36	51	91	118	100	62	67	25	11

*See Appendix G, Tally #2

EXPERIMENTAL PRETESTING YIELDS

Delta:	<8	8	9	10	11	12	13	14	15	≥16.
Avg # used in 4 Final Forms:	20	24	36	28	32	44	44	44	23	9
Avg yield/ Exp.Pretest: (All Item Types)**	1.0	1.3	1.9	3.4	4.1	7.3	6.5	4.3	2.3	0.8

Experimental Pretest Yields	<8	8	9	10	11	12	13	14	15	≥16
1 Pretest	1	1	2	3	4	7	7	4	2	1
2 Pretests	2	3	4	7	8	15	13	9	5	2
3 Pretests	3	4	6	10	12	22	20	13	7	2
4 Pretests	4	5	8	13	16	30	27	18	10	3
5 Pretests	5	6	10	17	20	37	34	22	12	4
6 Pretests	6	8	11	20	25	45	40	27	14	5
7 Pretests	7	9	13	24	29	52	47	31	17	6
8 Pretests	8	10	15	27	33	60	54	35	19	7
9 Pretests	9	11	17	30	37	67	61	40	22	7
10 Pretests	10	13	19	34	41	74	67	44	24	8
11 Pretests	11	14	21	37	45	82	74	49	26	9
12 Pretests	12	15	23	40	49	89	81	53	29	10
13 Pretests	13	17	25	44	53	97	87	57	31	11
14 Pretests	14	18	27	47	57	104	94	62	34	12
15 Pretests	15	19	29	50	61	112	101	66	36	12
16 Pretests	16	20	31	54	65	119	108	71	38	13
17 Pretests	17	22	32	57	70	127	114	75	41	14
18 Pretests	18	23	34	61	74	134	121	80	43	15
19 Pretests	19	24	36	64	78	141	128	84	46	16
20 Pretests	20	25	38	67	82	149	135	88	48	17

**Aggregation of data from Appendix F and Appendix G, Tally #2 for RCMP and SNCP

COST TO ATTAIN REQUISITE NUMBER OF ITEMS FOR FOUR FINAL FORMS

N_s = # of standard pretests
N_e = # of experimental pretests
V_e = $value of experimental pretest items

Cost = (N_s x $20,000[$/std. pretest]) + (N_e x $19,150[$/exp. pretest]) + (N_e x V_e)

Pretested items:	**V_e = $0**	**V_e = $10,985**
0 Exp.Prtst + 20 Std.Prtst =	$400,000	
1 Exp.Prtst + 18 Std.Prtst =	$379,150	$390,135
2 Exp.Prtst + 16 Std.Prtst =	$358,300	$380,269
3 Exp.Prtst + 14 Std.Prtst =	$337,450	$370,404
4 Exp.Prtst + 13 Std.Prtst =	$336,600	$380,538
5 Exp.Prtst + 12 Std.Prtst =	$335,750	$390,673
6 Exp.Prtst + 11 Std.Prtst =	$334,900	$400,807

54020-07635 • Y43M.7 • 209141